A Novel Approach for Joint Analytical and ML-assisted GSNR Estimation in Flexible Optical Network



(Article begins on next page)

# A Novel Approach for Joint Analytical and ML-assisted GSNR Estimation in Flexible Optical Network

F. Arpanaei(1), B. Shariati(2), P. Safari(2), M. Ranjbar Zefreh(3), J. A. Hernandez(1), A. Carena(4), J. Fischer(2), D. Larrabeiti(1)

(1) Universidad Carlos III de Madrid, Departamento de Ingeniería Telemática, farhad.arpanaei@uc3m.es

(2) Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute,

(3) Cisco Systems Italy S.r.l,

(4) Politecnico di Torino, Dipartimento di Elettronica e Telecomunicazioni.

**Abstract** *We propose a novel approach to perform QoT estimation relying on joint exploitation of machine learning and analytical formula that offers accurate estimation when applied to scenarios with heterogeneous span profiles and sparsely occupied links. Our approach significantly outperforms the widely used lightpath-level QoT estimation. ©2022 The Author(s)*

## Introduction

Accurate and real-time estimation of quality of transmission (QoT) can significantly contribute to the realization of low-margin optical networks [1,2]. The time-consuming estimators, i.e., split-step Fourier method (SSFM) [3], integral-based enhanced Gaussian noise (EGN) [4], and integral-based GN models applied to the link-level analysis [5], are not suitable for real-time QoT-aware network planning. However, closed-form models (CFMs) have been proposed that low offer complexity at the expense of reduced accuracy [6]. The inaccuracy may come from uncertainties of the input parameters or assumptions such as transparency and homogeneity of the spans [7-10].

The homogeneous characteristics of the span could be regarding their length, attenuation, dispersions, or non-linearities coefficients [8, 9]. On the other hand, the assumption of fully loaded links in the distance adaptive network planning reduces the CFMs' run time. However, this results in overestimating non-linear interferences (NLIs), which could increase the SNR design margin [11]. Machine learning (ML) based approaches have received significant attention targeting accuracy improvement while reducing the complexity [2, 12]. Most of the works in the literature focus on the estimation of end-to-end QoT of the lightpath (LP) under test, and a small number of them focus on refining the input parameters of the GN-based models to improve the accuracy while keeping the complexity level low [13-21].

In this paper, we propose a joint analytical and ML-assisted (JAM) model in which the exact values of the input parameters can be applied to the modified CFMs to perform QoT estimation. We perform a comprehensive numerical analysis and compare our proposal with the state-of-the-art approaches that justify our proposal's advantages for accurate yet fast QoT estimation.

## System Model and Analytical Formula

We considered a flexible optical network (FON), where the data plane mainly comprises two terminal nodes equipped with open reconfigurable add/drop multiplexers (ROADMs), pre-amplifiers, and boosters and intermediate nodes equipped with In-line amplifiers. We assume the modulation format level (MFL) of sliceable bandwidth variable transponders (S-BVTs) is selected based on the desired LP distance reach. A super-channel is formed by Nyquist shaped sub-channels (SbChs) having same symbol-rate and MFL. Thus, the occupied frequency slots by the SbChs are equal. In contrary to previous works [13-21], we apply a generalized SNR (GSNR) analytic model with MFL and long-haul LPs correction terms. The GSNR of SbCh $m$ on $r^{th}$ LP and the noise power of NLIs ($P_{\mathrm{NLI}}^{m,l,s}$) are obtained from Eq.(1)-Eq.(3).

$$GSNR_{\mathrm{LP}}^{m,r} \approx \left( \sum_{l=1}^{L^r} \sum_{s=1}^{S^{r,l}} \frac{1}{GSNR^{m,l,s}} \right)^{-1} - \Pi \qquad (1)$$

$$P_{\mathrm{NLI}}^{m,l,s} \approx \frac{8}{27\pi} B_{\mathrm{ch}}^{m} G_{\mathrm{ch}}^{m,l,s} \frac{(\gamma^{m,l,s})^2 L_{eff}^{l,s}}{|\beta_2^{l,s}|} \times \sum_{f'=1}^{N_{\mathrm{ch}}^{l}} (G_{\mathrm{ch}}^{n,l.s})^2 \Psi^{m,n,l,s}, \qquad (2)$$

$$\Psi^{n,m,l,s}|_{m\neq n} \approx \sinh^{-1}\left( \frac{\pi^2 |\beta_2^{l,s}| B_{\mathrm{ch}}^{m}}{2\alpha^{l,s}} \left[ f^n - f^m + \frac{B_{\mathrm{ch}}^{n}}{2} \right] \right) - \sinh^{-1}\left( \frac{\pi^2 |\beta_2^{l,s}| B_{\mathrm{ch}}^{m}}{2\alpha^{l,s}} \left[ f^n - f^m - \frac{B_{\mathrm{ch}}^{n}}{2} \right] \right) - \frac{5 R^n \Phi^{n,l,s} L_{\mathrm{eff}}^{l,s}}{3|f^n - f^m| L_{\mathrm{span}}^{l,s}}, \quad \Psi^{n,m,l,s}|_{m=n} \approx \sinh^{-1}\left( \frac{\pi^2 |\beta_2^{l,s}| (B_{\mathrm{ch}}^{m})^2}{4\alpha^{l,s}} \right). \qquad (3)$$

where $m, n, r, l,$ and $s$ are the indices of $m^{th}$ and $n^{th}$ SbCh, $r^{th}$ LP, $l^{th}$ link, and $s^{th}$ span, respectively. $L^r$ and $S^{r,l}$ are the number of links and spans of link $l$ for LP $r$, respectively. $\Pi = \frac{\sum_{l=1}^{L^r}\sum_{s=1}^{S^{r,l}} P_{\text{NLI}}^{m,l,s}}{\sum_{l=1}^{L^r}\sum_{s=1}^{S^r} P_{\text{ASE}}^{m,l,s} + \sum_{l=1}^{L^r}\sum_{s=1}^{S^r} P_{\text{NLI}}^{m,l,s}}$ is long-haul LPs correction term [22]. Moreover, $GSNR^{m,r,l,s} \approx \frac{P_{\text{ch}}^{m,l,s}}{P_{\text{ASE}}^{m,l,s}+P_{\text{NLI}}^{m,l,s}}$, and the noise power of erbium-doped fibre amplifiers is $P_{\text{ASE}}^{m,l,s} = hf^m N_{\text{F}}^{l,s}\left(G_{\text{span}}^{l,s} - 1\right)$, where $G_{\text{span}}^{l,s}$ is the amplifier gain of span $s$ on link $l$. Since the focus of this work is on comparing the ML-assisted GSNR estimators in span, link, and LP levels, we assume that the loss of each span is precisely compensated for by the span amplifier ($G_{\text{span}}^{l,s} = e^{2\alpha^{l,s} L_{\text{span}}^{l,s}}$). Indeed, the span length may be different. The CFM in Eq.(1) is a modified version of the incoherent CFM of equation 7.32 in [23]. The proposed GN with MFL correction term (GNWM) CFM is applicable for an LP with sparsely occupied links and heterogonous spans in a FON. In Eq.(1)-(3), the input parameters are the SbCh launch power ($P_{\text{ch}}^{m,l,s}$), the number of channels in link $l$ ($N_{\text{ch}}^{l}$), the fibre field loss ($\alpha^{l,s}$), the dispersion ($\beta_2^{l,s}$), and the fiber non-linearity ($\gamma^{l,s}$) coefficients, SbCh power density distribution ($G_{\text{ch}}^{m(n),l,s}$), bandwidth ($B_{\text{ch}}^{m(n)}$), and frequency center ($f^{m(n)}$) on channel $m(n)$, MFL correction factor ($\Phi^{n,l,s}$), symbol rate of SbChs ($R^{n,l,s}$), and amplifier's noise figure ( $N_{\text{F}}^{l,s}$). Finally, $L_{\text{span}}^{l,s}$ and , $L_{\text{eff}}^{l,s} = \frac{1-e^{-2\alpha^{l,s}.L_{\text{span}}^{l,s}}}{e^{2\alpha^{l,s}}}$ are the span and effective span length, respectively.

**Problem Formulation**

We developed two approaches to predict the GSNR of an LP. A first ML-based approach. A second JAM proposal: the link-level and span-level GSNRs are predicted according to an ML model, then we apply an accurate analytical model to concatenate the predicted link-level or span-level GSNRs to estimate the corresponding LPs' GSNRs of that links or spans [24].

For the ML approach, we randomly generated 10,000 LP samples (see Fig.1) as described below. Each sample consists of 12 attributes of an LP considered in (Eq.(1)-(3)). To generate the datasets (DSs), we randomly assigned values to those 12 attributes (in a list structure), assuring that they remain in the practical range of the relevant parameters. We specifically focus on span length heterogeneity and sparsely occupied links. Consequently, we defined the channel loading factor, which indicates the proportion of busy channels on each link. We consider the channel loading factor in the range of 10%-100%. Then, in the Ground Truth (GT) step, we apply the analytic model to estimate the launch power and GSNR of each LP end-to-end and extract their link-level and span-level GSNR according to a bit error rate (BER) threshold considering the MFL of the LPs. In this regard, the power of each channel is calculated according to the LOGON approach [5]. Next, the numbers of links and spans (in the last link) are calculated. Thus, the list of the channel's launch power, number of spans, and number of links are updated for each LP. Now, the LP-level GT (LP-GT) is constructed. Moreover, according to the calculated GSNR of each link and span related to the LPs, the Link-level GT (LL-GT) and Span-level GT (SL-GT) are formed. The final GT datasets of LP, link, and Span levels have 9526, 41590, and 196495 samples and are publicly available at [25]. We have some missing data in LPs because of the BER condition. We split LP-level GT into two sections, including Train/Cross-Validation and Test, with 80% and 20% of LP-GT samples. Then, LP-level, Link-level, and Span-level regressors are obtained by applying the ML models on Train/Cross-Validation LP-, LL-, and SL-GT, respectively. The predicted GSNRs of Test GT's LPs are obtained in the pure ML approach by applying an LP regressor.

On the other hand, in the proposed JAM algorithm, we exploit Eq.(1) without considering Eq.(2) and Eq.(3). Indeed, GSNRs of Test GT's LPs are estimated by applying the predicted GSNRs of corresponding spans and links obtained by the SL and LL regressors. To do so, we substitute the related estimated GSNRs obtained from SL and LL regressors in Eq.(1).

**Results**

In this study, to generate initial LP samples, we assume the LPs' symbol rates and their bandwidths equal 64 GBaud and 75 GHz (6x12.5 GHz), respectively. Thus, $N_{\text{ch}}^{l,s}$ = 60 in C-band, i.e., $f^{m(n)} \in$ {191.61, …, 195.95} THz. $\alpha^{l,s}$, $\beta_2^{l,s}$, $\gamma^{l,s}$, $N_{\text{F}}^{l,s}$ are 0.21 dB/km, -21.45x10$^{-27}$ s$^2$/m, 1.31x10$^{-3}$ (W.m)$^{-1}$, 6 dB, respectively. Additionally, $P_{\text{ch}}^{m,l,s} \in$ [-5,5] dBm with 0.01 dBm resolution and $\Phi^{n,l,s}$= {1,1,0.66,0.68,0.69, 0.62} is related to the PM- BPSK, -QPSK, -8QAM, -16QAM, -32QAM, and -64QAM that show with 1, 2, 3, 4, 5, and 6, respectively. Moreover, $L_{\text{span}}^{r,l,s}$, $N_{\text{span}}^{r,l}$, are in [50,120] km, and [1,10], respectively, and $N_{\text{Link}}^{r}$ equals 20. $BER_{threshold}$ = 3.8x10$^{-3}$ suitable for a 28% forward error correction overhead [26]. Thus, according to equations in [26], the GSNR thresholds for MFLs : 1-6 are 5.52, 8.53, 12.51, 15.19, 18.19, and 21.12 dB, respectively. Note the number of spans ($N_{\text{span}}^{r,l}$) and links ($N_{\text{Link}}^{r}$) are large enough in the initial raw data, and we refine them in the GT generation.

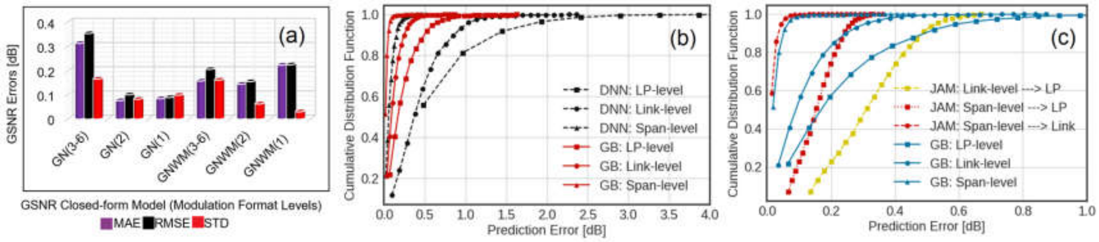


**Fig. 1:** (a) GSNR errors comparison of GN and GNWM (GN with modulation format level correction term) with EGN, (b) CDF of GB and DNN of span, link, and LP levels, (c) CDF of GB and JAM of span, link, and LP levels,

**Tab. 1:** Accuracy results for span (a),link (b), and LP (c) levels (a, b, c). Gradient Boosting (GB), Deep Neural Network (DNN).

| Model | Train/Cross-validation (80%, k=5) | | | Test (20%) | | |
|---|---|---|---|---|---|---|
| Merit | RMSE | MAE | $R^2$ (%) | RMSE | MAE | $R^2$(%) |
| GB | (1, 22, 87)x $10^{-3}$ | (23, 112, 221)x$10^{-3}$ | (99,99,99) | (34, 147, 307)x $10^{-3}$ | (24, 114, 224)x $10^{-3}$ | (99,99,99) |
| DNN | (9, 252, 747)x $10^{-3}$ | (71, 375, 980)x$10^{-3}$ | (99,98,98) | (31, 390, 636)x $10^{-3}$ | (997, 984, 979)x $10^{-3}$ | (99,98,97) |

We applied Gradient Boosting (GB) and Deep Neural Network (DNN) models to predict the GSNRs in three levels. In addition, to find the optimum hyperparameters, we apply the k-fold cross-validation technique with k = 5 and the grid search method. To validate the proposed analytic model (GNWM), we compared the GSNRs of the SbChs in randomly selected links obtained by applying the GNWM, GN, and the integral-based EGN model based on [4]. Since the running time of GSNR calculation using the EGN model for LPs with large number of spans is time-consuming, we considered 500 link-level samples with MFL: 3-6 and 5 samples for each MFL: 1 and 2. Fig.1(a) shows that the GN model for MFL: 1 and 2 is more accurate than GNWM. On the contrary, the GNWM is more accurate than GN for LPs with a shorter distance, i.e., MFL: 3-6. Thus, to generate the DSs, we applied GN for LPs with MFL:1 and 2 and GNWM for ones with MFL: 3-6. Additionally, the cumulative distribution functions (CDFs) curves in Fig1.(b) present the GSNR errors using GB ML and the DNN model in LP-, link-, and span-level. As shown in Fig1.(b), about 99% of the prediction errors (GT GSNRs - Estimated GSNRs) of SL, LL, and LP levels are lower than 0.1, 0.4, and 0.8 dB for GB and 0.3, 1.1, and 1.95 dB for DNN, respectively. Moreover, in Fig2. (c), CDFs' curves of estimated GSNR errors of LPs in Test GT obtained by applying the corresponding estimated GSNRs of spans, and links, to Eq.(1) shows with JAM: Span-level ---> LP and JAM: Link-level ---> LP, respectively. Also, curve JAM: Span-level ---> link shows the estimated LLs' Test GSNRs by applying corresponding estimated span GSNRs, to Eq.(1). The results show that about 99% of the prediction errors of LPs' GSNRs reduce from 0.8 dB with applying GB to 0.3 and 0.58 dB for JAM: Span-level---> LP and JAM: Link-level--->LP, respectively. This value for link-level GSNRs improves from 0.4 dB using GB to 0.09 dB using the JAM: Span-level --> link. A chart of the GB's results, highlighting the promising results of the JAM model, is shown in Fig 1. (b) and (c). RMSE and MAE are reported in Tab.1 and 2 for GB, DNN, and JAM. The results confirm CDFs' behavior and emphasize that the GB is more accurate than DNN regarding the synthesized DSs. However, the GSNR calculation for an LP averagely took longs in order 130 and 200 μsec, respectively, from the link and span levels GSNRs (JAM approach). The run time of an LP's GSNR calculation is about in order 14 msec for the proposed CFM.

**Tab. 2:** Accuracy results for joint analytical and ML (JAM)

| Model | Link-level --> LP | Span-level --> LP | Span-level -->Link |
|---|---|---|---|
| RMSE | 11 x$10^{-3}$ | 24 x$10^{-3}$ | 49 x$10^{-3}$ |
| MAE | 31 x$10^{-3}$ | 15 x$10^{-3}$ | 15 x$10^{-3}$ |

## Conclusion

The results show that for LPs with heterogeneous span profiles and sparsely occupied links our joint analytical and ML (JAM) proposal outperforms pure ML models and the proposed GN with MFL correction term in terms of accuracy and speed, respectively. Indeed, the LPs' GSNRs estimation accuracy improves by 0.4 dB using JAM in comparison to pure ML. Additionally, the speed of LPs' GSNRs calculation using JAM is in microseconds, whereas using analytical model it is in milliseconds.

## Acknowledgements

Farhad Arpanaei acknowledges support from the CONEX-Plus programme funded by Universidad Carlos III de Madrid and the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 801538. The authors would like to acknowledge the support of the EU-funded B5G-OPEN project (grant no. 101016663) and the Spanish projects ACHILLES (PID2019-104207RB-I00) and Go2Edge (RED2018-102585-T). This work was supported by the Italian Ministry for University and Research (PRIN 2017, project FIRST).

## Reference


[1] Y. Pointurier, "Design of low-margin optical networks," in Journal of Optical *Communications and Networking*, vol. 9, no. 1, pp. A9-A17, Jan. 2017, DOI: 10.1364/JOCN.9.0000A9.

[2] Y.Pointurier, "Machine learning techniques for quality of transmission estimation in optical networks," J. Opt. Commun. Netw. 13, B60-B71 (2021).

[3] J. Shao, X. Liang and S. Kumar, "Comparison of Split-Step Fourier Schemes for Simulating Fiber Optic Communication Systems," in *IEEE Photonics Journal*, vol. 6, no. 4, pp. 1-15, Aug. 2014, Art no. 7200515, DOI: 10.1109/JPHOT.2014.2340993.

[4] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "EGN model of non-linear fiber propagation," Opt. Express 22, 16335-16362 (2014).

[5] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang and F. Forghieri, "The GN-Model of Fiber Non-Linear Propagation and its Applications," in *Journal of Lightwave Technology*, vol. 32, no. 4, pp. 694-721, Feb.15, 2014, DOI: 10.1109/JLT.2013.2295208.

[6] M. Ranjbar Zefreh, F. Forghieri, S. Piciaccia and P. Poggiolini, "Accurate Closed-Form Real-Time EGN Model Formula Leveraging Machine-Learning Over 8500 Thoroughly Randomized Full C-Band Systems," in Journal of Lightwave Technology, vol. 38, no. 18, pp. 4987-4999, 15 Sept.15, 2020, DOI: 10.1109/JLT.2020.2997395.

[7] E. Seve, J. Pesic, C. Delezoide, S. Bigo and Y. Pointurier, "Learning process for reducing uncertainties on network parameters and design margins," in *Journal of Optical Communications and Networking*, vol. 10, no. 2, pp. A298-A306, Feb. 2018, DOI: 10.1364/JOCN.10.00A298.

[8] J. Pesic, M. Lonardi, N. Rossi, T. Zami, E. Seve and Y. Pointurier, "How Uncertainty on the Fiber Span Lengths Influences QoT Estimation using Machine Learning in WDM Networks," *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, 2020, pp. 1-3.

[9] J. Pesic, "Missing Pieces Currently Preventing Effective Application of Machine Learning to QoT Estimation in the Field," *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, 2021, pp. 1-3.

[10] M. Lonardi, J. Pesic, T. Zami, E. Seve and N. Rossi, "Machine learning for quality of transmission: a picture of the benefits fairness when planning WDM networks," in *Journal of Optical Communications and Networking*, vol. 13, no. 12, pp. 331-346, December 2021, DOI: 10.1364/JOCN.433412.

[11] N. Morette, I. F. de Jauregui Ruiz, H. Hafermann and Y. Pointurier, "On the Robustness of a ML-based Method for QoT Tool Parameter Refinement in Partially Loaded Networks," *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, 2022, pp. 1-3.

[12] Lu Zhang, Xin Li, Ying Tang, Jingjie Xin, Shanguo Huang, A survey on QoT prediction using machine learning in optical networks, Optical Fiber Technology, Volume 68, 2022, 102804, ISSN 10685200, DOI: https://doi.org/10.1016/j.yofte.2021.102804.

[13] A. D'Amico *et al.*, "Using machine learning in an open optical line system controller," in *Journal of Optical Communications and Networking*, vol. 12, no. 6, pp. C1-C11, June 2020, DOI: 10.1364/JOCN.382557.

[14] E. Seve, J. Pesic and Y. Pointurier, "Associating machine-learning and analytical models for quality of transmission estimation: combining the best of both worlds," in *Journal of Optical Communications and Networking*, vol. 13, no. 6, pp. C21-C30, June 2021, DOI: 10.1364/JOCN.411979.

[15] D. Uzunidis, A. Stavdas, P. Kasnesis, C. Patrikakis and A. Lord, "Enhancing Closed-Form Based Physical Layer Performance Estimations in EONs Via Machine Learning Techniques," *2021 European Conference on Optical Communication (ECOC)*, 2021, pp. 1-4, DOI: 10.1109/ECOC52684.2021.9605893.

[16] H. Lun *et al.*, "Machine-learning-based telemetry for monitoring long-haul optical transmission impairments: methodologies and challenges [Invited]," in *Journal of Optical Communications and Networking*, vol. 13, no. 10, pp. E94-E108, October 2021, DOI: 10.1364/JOCN.426826.

[17] J. Müller *et al.*, "Estimating Quality of Transmission in a Live Production Network using Machine Learning," *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, 2021, pp. 1-3.

[18] J. Müller, S. K. Patri, T. Fehenberger, C. Mas-Machuca, H. Griesser and J. -P. Elbers, "A QoT Estimation Method using EGN-assisted Machine Learning for Network Planning Applications," *2021 European Conference on Optical Communication (ECOC)*, 2021, pp. 1-4, DOI: 10.1109/ECOC52684.2021.9606064.

[19] M. Ibrahimi *et al.*, "Machine learning regression for QoT estimation of unestablished lightpaths," in *Journal of Optical Communications and Networking*, vol. 13, no. 4, pp. B92-B101, April 2021, DOI: 10.1364/JOCN.410694.

[20] S. Allogba, S. Aladin and C. Tremblay, "Machine-Learning-Based Lightpath QoT Estimation and Forecasting," in Journal of Lightwave Technology, 2022, DOI: 10.1109/JLT.2022.3160379.

[21] G. Bergk, B. Shariati, P. Safari, J. Fischer, "ML-assisted QoT estimation: a dataset collection and data visualization for dataset quality evaluation," JOCN, vol. 14, no. 3, pp. 43-55, Mar 2022.

[22] M. R. Zefreh, F. Forghieri, S. Piciaccia and P. Poggiolini, "Nonlinearity Assessment in Long Haul Dispersion Managed Fiber Optic Links," *2020 Italian Conference on Optics and Photonics (ICOP)*, 2020, pp. 1-3, DOI: 10.1109/ICOP49690.2020.9300336.

[23] P. Poggiolini, Y. Jiang, A. Carena and F. Forghieri, "Analytical modeling of the impact of fiber non-linear propagation on coherent systems and networks" in Enabling Technologies for High Spectral-Efficiency Coherent Optical Communication Networks, New York, NY, USA:Wiley, pp. 247-310, 2016.

[24] K. Kaeva, J. Myyry, K. Grobe, H. Grießer and G. Jervan, "Concatenated GSNR Profiles for End-to-End Performance Estimations in Disaggregated Networks," 2022 Optical Fiber Communications Conference and Exhibition (OFC), 2022, pp. 1-3.

[25] Farhad Arpanaei, "Data Sets for SNR Estimation in Flexible Optical Networks: Lightpath, Link, and Span Levels," (2022). DOI: 10.5281/zenodo.6529744.

[26] B. Shariati, A. Mastropaolo, N. Diamantopoulos, J. M. Rivas-Moscoso, D. Klonidis and I. Tomkos, "Physical-layer-aware performance evaluation of SDM networks based on SMF bundles, MCFs, and FMFs," in Journal of Optical Communications and Networking, vol. 10, no. 9, pp. 712-722, Sept. 2018, DOI: 10.1364/JOCN.10.000712.